

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 1, 2001, 16:20:32 ; Search time 17.21 Seconds

(without alignments)  
123.847 Million cell updates/sec

Title: US-09-331-631a-24\_COPY\_29\_94

Perfect score: 382  
Sequence: 1 HDEDDRGGHSLQDCVQRC.....EQEEQGRGRWHGEGERE 66

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 88757 seqs, 32294092 residues

Total number of hits satisfying chosen parameters: 88757

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_39:\*

Pred. No. is the number of result's predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	133	34.8	573	1	GLB1_MAIZE
2	90	23.6	588	1	WCLB_GOSHI
3	85	22.3	639	1	GLCA_SOYBN
4	83.5	21.9	605	1	WCLA_GOSHI
5	76.5	20.0	338	1	LEGB_PEA
6	73	19.1	605	1	GLCA_SOYBN
7	72	18.8	852	1	SRCH_RABIT
8	70.5	18.5	185	1	T2_MOUSE
9	70	18.3	1130	1	REPT_MOUSE
10	70	18.3	1224	1	PER_DROME
11	70	18.3	1549	1	TRHY_SHEEP
12	69.5	18.2	336	1	FLA_MOUSE
13	69.5	18.2	1208	1	PER_DROYA
14	69	18.1	661	1	PER_DROSE
15	68.5	17.9	654	1	CIK4_MOUSE
16	68.5	17.9	655	1	CIK4_RAT
17	68	17.8	654	1	CIK4_MOUSE
18	67.5	17.7	1394	1	CNGA_BOVIN
19	67	17.5	228	1	DH2_ORYSA
20	67	17.5	516	1	LEGB_GOSHI
21	67	17.5	653	1	CIK4_HUMAN
22	67	17.5	1121	1	DDX8_ARATH
23	66.5	17.4	656	1	PER_DROSI
24	66.5	17.4	676	1	PER_DROMA
25	66	17.3	407	1	IE68_HSVSA
26	66	17.3	1407	1	TRHY_RABIT
27	65.5	17.1	412	1	YNP1_CAHEL
28	65.5	17.1	1898	1	TRHY_HUMAN
29	65	17.0	355	1	SRP_MUSSP
30	65	17.0	524	1	SRP_SOYBN
31	65	17.0	888	1	YCB4_YEAST
32	65	17.0	1063	1	SPT5_YEAST
33	65	17.0	2248	1	CYAL_DROME

34	64.5	16.9	242	1	THYL_HUMAN	P20396	homo sapien
35	64.5	16.9	355	1	CPDI_DROME	P22058	drosophila
36	64	16.8	493	1	11S3_HELAN	P19084	heiliantus
37	64	16.8	543	1	P2CG_BOVIN	P79126	bos taurus
38	64	16.8	681	1	VE1_HPV10	P36720	human papil
39	64	16.8	797	1	VG48_HSVSA	001033	herpesvirus
40	63.5	16.6	708	1	GBF_DICDI	P36417	dictyosteli
41	63.5	16.6	905	1	SNE5_YEAST	P18480	saccharomyc
42	63.5	16.6	1021	1	MANA_RHOWR	P49425	rhodothermu
43	63	16.5	699	1	SRCH_HUMAN	P23327	homo sapien
44	62	16.2	178	1	2SS2_BRANA	P01090	brassica na
45	62	16.2	258	1	2SS2_RICCO	P01089	ricinus com

## ALIGNMENTS

```
RESULT 1
GLB1_MAIZE STANDARD: PRT: 573 AA.
AC P15590:
DT 01-APR-1990 (Rel. 14, Created)
DT 01-AUG-1990 (Rel. 15, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE GLOBULIN-1 S ALLELE PRECURSOR (GLB1-S) (7S-Like).
GN GLB1.
OS Zea mays (Maize).
OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
NC Magnoliophyta; Liliopsida; Poales; Poaceae; Zea.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-CV. INBRED LINE VA26;
RA Belanger F.C., Kriz A.L.;
RT "Molecular characterization of the major maize embryo globulin encoded
RT by the Glb1 gene."
RL Plant Physiol. 91:636-643(1989).
RN [2]
RP SEQUENCE OF 87-100.
RX MEDLINE=89374022; PubMed=2775172;
RA Kriz A.L.;
RT "Characterization of embryo globulins encoded by the maize Glb
RT genes."
RL Biochem. Genet. 27:239-251(1989).
CC -I- PPM: THREE PROTEIN-PROCESSING STEPS OCCUR IN THE FORMATION OF THE
CC MATURE PROTEIN FROM THE PRIMARY TRANSLATION PRODUCT.
CC -I- POLYMORPHISM: THE THREE MOST COMMONLY OCCURRING GLB1 ALLELES HAVE
CC THE DESIGNATION L, I, AND S FOR LARGE, INTERMEDIATE, AND SMALL
CC PROTEINS, RESPECTIVELY.
CC -I- SIMILARITY: TO OTHER 7S SEED STORAGE PROTEINS (PHASEOLIN, VICILIN,
CC CONVICILIN, CONGLYCININ, ETC.).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL: M24845; AAA3467.1; -.
DR HSSP: P50477; ICAM.
DR MAITZEDB: 30181; -.
DR INTERPRO: IPR001113; -.
DR PFAM: PF00546; Seedstore_7s; 1.
KW Seed storage protein; Signal.
FT SIGNAL 1 18
FT PROPEP 19 86
FT CHAIN 87 573 GLOBULIN-1 S ALLELE.
FT CARBOHYD 349 349 N-LINKED (GLICAC...) (POTENTIAL).
SQ SEQUENCE 573 AA; 65029 MW; 525ED1D0A062976 CRC64;
```

Query Match

34.8%; Score 133; DB 1; Length 573;

```

Best Local Similarity 36.2%; Pred. No. 8,5e+07; Indels 6; Gaps 3;
Matches 25; Conservative 15; Mismatches 27; Indels 6; Gaps 3;

OY      2 DDEDDRGGSHLQOCVORCRDPRPYSHAACVOECRDDO----QQHGRHEDEEGRGNG 57
          ||| :|||::||| : |||::|||::||| : |||::|||
DB      26 DDNNHHNGGKSGRCVRRC-EDRPWHQRPRCLCECFEEREREKRORSRHEADRSGEGSS 84
          ::|||::|||::|||::|||::|||::|||::|||::|||::|||

OY      58 WHGEGEREE 66
          | ||||
DB      85 -EDERERRRO 92

RESULT  2
VCUB_GOSHI STANDARD; PRT; 588 AA.
ID VCUB_GOSHI
AC P09801;
DT 01-MAR-1989 (Rel. 10, Created)
DT 01-MAR-1989 (Rel. 10, Last sequence update)
DI 15-JUL-1999 (Rel. 38, Last annotation update)
DE VICLIN C72 PRECURSOR (ALPHA-GLOBULIN B).
OS Gossypium hirsutum (Upland cotton).
OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;
OC Malvales; Malvaceae; Gossypium.
RN [1]
RP SEQUENCE FROM N.A.
RA Chlan C.A., Pyle J.B., Legocki A.B., Dure L. III;
RT "Developmental biochemistry of cottonseed embryogenesis and
RT germination. XVIII. cDNA and amino acid sequences of the members of
RT the storage protein families."
RL Plant Mol. Biol. 7:475-489(1986).
CC CC
CC -! FUNCTION: SEED STORAGE PROTEIN.
CC -! SUBCELLULAR LOCATION: CYTOLEONARY MEMBRANE-BOUND VACUOLAR PROTEIN
CC BOODIES.
CC -! SIMILARITY: TO OTHER 7S SEED STORAGE PROTEINS (PHASEOLIN, VICILIN,
CC CONVICILIN, CONGLYCININ, ETC.).
-----
CC CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
CC CC EMBL; M16891; AAA33071.1; -.
CC DR PIR; A30838; FMCNAB.
CC DR HSSP; P50477; ICAX.
CC DR INTERPRO; IPR001113; -.
CC DR PFAM; PF00546; Seedstore_7s; 1.
CC DR Seed storage protein; Signal.
CC FT SIGNAL 1
CC FT CHAIN 26 588 VICILIN C72.
CC SQ SEQUENCE 588 AA; 69729 MW; 63E699B25A8BADBE CRC64;

Query Match 23.6%; Score 90; DB 1; Length 588;
Best Local Similarity 31.2%; Pred. No. 0.023;
Matches 20; Conservative 14; Mismatches 28; Indels 2; Gaps 2;

OY      3 DDDDRGGSHLQOCVORCRDPRPYSH-ARCVOECRDDQQH-GRHEDEEGRGGRGNHG 60
          ::|||::|||::|||::|||::|||::|||::|||::|||::|||
DB      114 EOEEOQSORFOECOOHCQOEOPREPKQCVCRCFCRRFYQENPMWRFEEREEAEETDEG 173
          ::|||::|||::|||::|||::|||::|||::|||::|||::|||

OY      61 EGER 64
          | |
DB      174 EOEQ 177

RESULT  3
GLCX_SOYBN STANDARD; PRT; 639 AA.
ID GLCX_SOYBN
DC P11827;
```

DT	01-OCT-1989	(Rel. 12, created)
DT	01-OCT-1989 <td>(Rel. 12, last sequence update)</td>	(Rel. 12, last sequence update)
DT	01-AUG-1992 <td>(Rel. 23, last annotation update)</td>	(Rel. 23, last annotation update)
DE	BETA-CONGLICLININ, ALPHA CHAIN PRECURSOR.	CG-1.
OS	Glycine max (Soybean).	
OC	Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;	
OC	Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids I;	
OC	Fabiales; Fabaceae; Papilionoideae; Glycine.	
RN	[1]	
RP	SEQUENCE FROM N.A.	
RX	MEDLINE=8625087; PubMed=3013879;	
RA	Doyle J.J., Schuler M.A., Godelle W.D., Zenger V., Beachy R.N.,	
RA	Slightom J.L.;	
RT	"the glycosylated seed storage proteins of Glycine max and Phaseolus	
RT	vulgaris. Structural homologies of genes and proteins.";	
RL	J. Biol. Chem. 261:9228-9238(1986).	
RN	[2]	
RP	SEQUENCE OF 340-639 FROM N.A.	
RX	MEDLINE=83143289; PubMed=6897678;	
RA	Schuler M.A., Ladin B.P., Pollaco J.C., Freyer G., Beachy R.N.;	
RT	"Structural sequences are conserved in the genes coding for the	
RT	alpha, alpha' and beta-subunits of the soybean 7S seed storage	
RT	protein.";	
RL	Nucleic Acids Res. 10:8245-8261(1982).	
CC	-I- FUNCTION: SEED STORAGE PROTEIN. ACCUMULATES DURING SEED	
CC	DEVELOPMENT AND IS HYDROLYZED AFTER GERMINATION TO PROVIDE A	
CC	CARBON AND NITROGEN SOURCE FOR THE DEVELOPING SEEDLING.	
CC	-I- SUBUNIT: THE ALPHA-, ALPHA', AND BETA-SUBUNIT'S ASSOCIATE IN	
CC	VARIOUS COMBINATIONS TO FORM TRIMERIC PROTEINS.	
CC	-I- SUBCELLULAR LOCATION: CYTOLEDONARY MEMBRANE-BOUND VACUOLAR PROTEIN	
CC	BODIES.	
CC	-I- SIMILARITY: TO OTHER 7S SEED STORAGE PROTEINS (PHASEOLIN, VICILIN,	
CC	CONVICILIN, CONLYCININ, ETC.).	
CC	-----	
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration	
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -	
CC	the European Bioinformatics Institute. There are no restrictions on its	
CC	use by non-profit institutions as long as its content is in no way	
CC	modified and this statement is not removed. Usage by and for commercial	
CC	entities requires a license agreement (see <a href="http://www.isb-sib.ch/announce/">http://www.isb-sib.ch/announce/</a>	
CC	or send an email to <a href="mailto:license@isb-sib.ch">license@isb-sib.ch</a> ).	
CC	-----	
DR	EMBL; M13759; AB01374.1; -	
DR	EMBL; J01290; -; NOT_ANNOTATED_CDS.	
DR	PIR; B24810; B24810.	
DR	HSSP; P02853; 2PRL.	
DR	INTERPRO; IPR001113; -	
DR	FRAM; PF00546; Seedstore_7s; 2.	
KW	Seed storage protein; Signal; Glycoprotein; Multigene family.	
FT	SIGNAL	1 22
FT	PROPEP	23 62
FT	CHAIN	63 639 BETA-CONGLYCININ, ALPHA' CHAIN.
FT	CARBOHYD	277 277 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	551 551 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CONFLICT	543 543 P -> L (IN REF. 2).
FT	CONFLICT	549 549 M -> V (IN REF. 2).
FT	CONFLICT	608 608 S -> T (IN REF. 2).
SO	SEQUENCE	639 AA; 74325 MW; 469BF24C79651E3F CRC64;
Query Match	22.3%; Score 85; DB 1; Length 639;	
Best Local Similarity	24.3%; Pred. No. 0.081;	
Matches	26; Conservative 9; Mismatches 20; Indels 52; Gaps 5;	
OY	12 SLQGVQRCROEPRRY-----SHARC-----VDECRDQ-----QONGHNE 47	
DB	34 SHNKLCRSCSEKDSYRNQACNARCNLLKVEEECEGEGQIAPRRQNREREONGSKE 93	
OY	48 QEE-----EGQGRGSH-----GEGREE 66	
DB	94 EDEGGRRRRRRRRPRQDNEENHOKENENHNRKEENKGGSGEE 140	

```

RESULT 4
VCLA_GOSHI
ID VCLA_GOSHI STANDARD: PRI: 605 AA.
AC P09799;
DT 01-MAR-1989 (Rel. 10, Created)
DT 01-MAR-1989 (Rel. 10, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE VICILIN GC72-A.PRECURSOR (ALPHA-GLOBULIN A).
OS Gossypium hirsutum (Upland cotton).
OC Eukaryota: Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC Magnoliophyta: eudicotyledons: core eudicots; Rosidae; eurosids II;
OC Malvales; Malvaceae; Gossypium.
RN [1]
RP SEQUENCE FROM N.A.
RA Chlan C.A., Borroto K., Kamalay J.A., Dure L. III;
RT "Developmental Biochemistry of cottonseed embryogenesis and
RT germination. XIX. Sequences and genomic organization of the alpha
RT globulin (vicilin) genes of cottonseed.";
RT Plant Mol. Biol. 9:533-546(1987).
CC -1- FUNCTIONAL: SEED STORAGE PROTEIN.
CC -1- SUBCELLULAR LOCATION: CYTLEDONARY MEMBRANE-BOUND VACUOLAR PROTEIN
CC BODIES.
CC -1- SIMILARITY: TO OTHER 7S SEED STORAGE PROTEINS (PHASEOLIN, VICILIN,
CC CONVICILIN, CONGLYCININ, ETC.).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; M19378; AAA33069.1; -.
CC PIR; S06398; S06398.
CC DR HSSP; P50477; ICAX.
CC DR INTERPRO; IPR001113; -.
CC PRAM; PF00546; Seedstore_7s; 1.
CC KM Seed storage protein; Signal.
CC FT SIGNAL 1 23
CC FT CHAIN 24 605 VICILIN GC72-A.
CC SEQUENCE 605 AA; 71049 MW; C9DB9371C976953B CRC64;
QY
Db 1 HDEDDRRGSHSLQCVQRC--RQRPYSHARQVQECRDDQDQ-----HGRHDE 49
QY : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 110 YEKQOQOQDPKQFKQRCQRCQOMQEQRPKRC-QCYEKREQYQEQDPWKGERENKREFE 168
QY 50 EEOGGRGHWGGEREE 66
QY : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 169 EES-----DEGEQQQ 178
RESULT 5
LEGB_PEA
ID LEGB_PEA STANDARD: PRI: 338 AA.
AC P14594;
DT 01-JAN-1990 (Rel. 13, Created)
DT 01-JAN-1990 (Rel. 13, Last sequence update)
DT 01-AUG-1992 (Rel. 23, Last annotation update)
DE LEGUMIN B (FRAGMENT).
DE LEGB.
OS Pisum sativum (garden pea).
OC Eukaryota: Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC Magnoliophyta: eudicotyledons: core eudicots; Rosidae; eurosids I;
OC Fabales; Fabaceae; Papilionoideae; Pisum.
RN [1]
RP SEQUENCE FROM N.A.

```

[illegible]

CC VACUOLAR PROTEIN BODIES.  
CC -1- SIMILARITY: TO OTHER 7S SEED STORAGE PROTEINS (PHASEOLIN, VICILIN,  
CC CONVICILIN, CONGLYCININ, ETC.).  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL: X17698; CAA35691.1; -  
CC PIR: S14681; FWSYBA.  
CC HSSP: P50477; ICAM.  
CC INTERPRO: IPR001113; -  
CC PFAM: PF00546; Seedstore\_7s; 1.  
CC Seed storage protein; Signal; Glycoprotein; Multigene family.  
CC STGNL 1 22  
CC PROPEP 23 62  
CC CHAIN 63 605 BETA-CONGLYCININ, ALPHA CHAIN.  
CC CARBOHYD 261 261 N-LINKED (GLCNAC. . .) (POTENTIAL).  
CC CARBOHYD 517 517 N-LINKED (GLCNAC. . .) (POTENTIAL).  
CC SEQUENCE 605 AA; 70293 MW; CBEBA30506BBBC57 CRC64;  
SO  
  
Query Match 19.1%; Score 73; DB 1; Length 605;  
Best Local Similarity 30.3%; Pred. No. 1.3;  
Matches 20; Conservative 6; Mismatches 14; Indels 26; Gaps 3;  
  
OY 15 OCVCORCROERPRY-----SHARC-----VOEC-----RDQQQHGHRHQ 48  
DB 37 KCLQSCNSESDSYRONACHRCMLKVEKECEGEIIPRPPRQHPERPPQGEKEED 96  
OY 49 EEEQGR 54  
DB 97 EDEQPR 102  
  
RESULT 7  
SRCH\_RABIT  
ID SRCH\_RABIT STANDARD; PRT; 852 AA.  
AC P16230;  
DT 01-APR-1990 (Rel. 14, Created)  
DT 01-APR-1990 (Rel. 14, Last sequence update)  
DT 15-JUL-1999 (Rel. 38, Last annotation update)  
DE SARCOPLASMIC RETICULUM HISTIDINE-RICH CALCIUM-BINDING PROTEIN  
DE PRECURSOR (HCP).  
OS Oryctolagus cuniculus (Rabbit).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.  
RN [1]  
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.  
RC TISSUE-SKELETAL MUSCLE;  
RX MEDLINE-90036884; PubMed-2808365;  
RA Hofmann S.L., Goldstein J.L., Orth K., Moormaw C.R., Staughter C.A.,  
RA Brown M.S.;  
RT "Molecular cloning of a histidine-rich Ca2+-binding protein of  
RT sarcoplasmic reticulum that contains highly conserved repeated  
RT elements".  
RL J. Biol. Chem. 264:18083-18090(1989).  
CC -1- FUNCTION: HCP MAY PLAY A ROLE IN THE REGULATION OF CA(2+)  
CC SEQUESTRATION OR RELEASE IN THE SR OF SKELETAL AND CARDIAC  
CC MUSCLE.  
CC -1- SUBCELLULAR LOCATION: SARCOPLASMIC RETICULUM LUMEN.  
CC -1- SIMILARITY: STRONG TO HUMAN HRC.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL: X67863; CAA48048.1; -

CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch).  
CC EMBL: J05080; AAA31279.1; -  
CC PIR: A34373; A34373.  
CC INTERPRO: IPR002134; -  
CC PROSITE: PS00328; HCP; 10.  
CC Calcium-binding; Signal; Repeat.  
CC STGNL 1 27  
CC CHAIN 28 852  
CC MOD\_RES 28 28  
CC DOMAIN 59 100  
CC REPEAT 59 79  
CC REPEAT 80 100  
CC DOMAIN 199 470 10 X TANDEM REPEATS, ACIDIC.  
CC REPEAT 199 224  
CC REPEAT 225 253  
CC REPEAT 254 282  
CC REPEAT 283 310  
CC REPEAT 311 339  
CC REPEAT 340 367  
CC REPEAT 368 395  
CC REPEAT 396 423  
CC REPEAT 424 451  
CC REPEAT 452 470  
CC DOMAIN 471 585  
CC DOMAIN 721 733  
CC DOMAIN 780 826  
CC SEQUENCE 852 AA; 96117 MW; A430C83BE94B30 CRC64;  
SO  
  
Query Match 18.8%; Score 72; DB 1; Length 852;  
Best Local Similarity 27.8%; Pred. No. 2.3;  
Matches 20; Conservative 11; Mismatches 21; Indels 20; Gaps 3;  
  
OY 7 RRGSHLQOCVRCRCRERPRYSHARCVRDQDQGR-----HEDEEGGRGRG----- 57  
DB 663 KRGGSE-----EEEDQGTGHHSHLEDEDEEGHGRSLQSDQDEDDRGESAKVOA 714  
OY 58 ---WHGGEREE 66  
DB 715 PLRHHREREE 726  
  
RESULT 8  
T2\_MOUSE  
ID T2\_MOUSE STANDARD; PRT; 185 AA.  
AC Q06666;  
DT 01-FEB-1995 (Rel. 31, Created)  
DT 01-FEB-1995 (Rel. 31, Last sequence update)  
DT 01-NOV-1997 (Rel. 35, Last annotation update)  
DE OCTAPEPTIDE-REPEAT PROTEIN T2.  
GN SRST OR T2.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-BALB/C; TISSUE-MACROPHAGE;  
RX MEDLINE-93092084; PubMed-1458435;  
RA di Carlo M., Montana G., Romanino D.P., Monteleone D.;  
RA "A mouse repeat sequence conserved in eukaryotic genomes".  
RL J. Submicrosc. Cytol. Pathol. 24:467-472(1992).  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL: X67863; CAA48048.1; -

DR MGD: MGI:107677; 1st.  
 FT Repeat.  
 SO SEQUENCE 185 AA; 22805 MW; 8995BECAEC383971 CRC64;

Query Match 18.5%; Score 70.5; DB 1; Length 185;  
 Best Local Similarity 35.5%; Pred. No. 0.76;  
 Matches 22; Conservative 10; Mismatches 23; Indels 7; Gaps 3;

QY 6 DRGRHSIQGVQRCRO-ERPRYSHARCVCQCRDQOQHGEDEEGRGRMHGEEER 64  
 Db 120 ERGRREAR--QRCREARGRRA---EKGRGRGRGRVEAEGRGKMRQEA 173  
 QY 65 EE 66  
 Db 174 TE 175

RESULT 9  
 ID REPT\_MOUSE STANDARD; PRT; 1130 AA.  
 AC P97347;  
 DT 01-NOV-1997 (Rel. 35, Created)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DT 01-OCT-2000 (Rel. 40, Last annotation update)  
 DE REPERIN.  
 GN RPTN.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-NMRI: TISSUE-SKIN;  
 RL MEDLINE=97422611; PubMed=9268637;  
 RA Kriegl P., Schuppler M., Koesters R., Mincheva A., Lichter P.,  
 RA Marks F.;  
 RT "Repetin (Rptn), a new member of the 'tused gene' subgroup within the  
 RT S100 gene family encoding a murine epidermal differentiation  
 RT protein";  
 RL Genomics 43:339-348(1997).  
 CC -1- FUNCTION: NOVEL POTENTIAL PRECURSOR PROTEIN OF THE CORNIFIED CELL  
 CC ENVELOPE.  
 CC -1- SIMILARITY: IN THE N-TERMINAL SECTION; BELONGS TO THE S-100  
 CC FAMILY.  
 CC -1- SIMILARITY: CONTAINS 2 EF-HAND CALCIUM-BINDING DOMAINS.  
 CC -----  
 CC "THIS SWISS-PROT ENTRY IS COPYRIGHT. IT IS PRODUCED THROUGH A COLLABORATION  
 CC BETWEEN THE SWISS INSTITUTE OF BIOINFORMATICS AND THE EMBL OUTSTATION -  
 CC THE EUROPEAN BIOINFORMATICS INSTITUTE. THERE ARE NO RESTRICTIONS ON ITS  
 CC USE BY NON-PROFIT INSTITUTIONS AS LONG AS ITS CONTENT IS IN NO WAY  
 CC MODIFIED AND THIS STATEMENT IS NOT REMOVED. USAGE BY AND FOR COMMERCIAL  
 CC ENTITIES REQUIRES A LICENSE AGREEMENT (SEE <http://www.isb-sib.ch/announce/>  
 CC OR SEND AN EMAIL TO [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: X99251; CAA67624.1; -  
 DR HSSP: P02633; 3ICB.  
 DR MGD: MGI:1099055; RPTN.  
 DR INTERPRO: IPR001751; -  
 DR INTERPRO: IPR002048; -  
 DR PFAM: PF01023; S\_100; 1.  
 DR PFAM: PF00036; efhand; 1.  
 DR PROSITE: PS00018; EF\_HAND; 1.  
 DR PROSITE: PS00303; S100\_CABP; 1.  
 KM Repeat: Calcium-binding.  
 FT DOMAIN 1 91  
 FT CA\_BIND 22 33 S-100 LIKE.  
 FT CA\_BIND 62 73 SITE I (LOW AFFINITY) (POTENTIAL).  
 SO SEQUENCE 1130 AA; 129884 MW; 7D97BDD25151918D CRC64;

Query Match 18.3%; Score 70; DB 1; Length 1130;  
 Best Local Similarity 26.2%; Pred. No. 4.9;

Matches 17; Conservative 14; Mismatches 28; Indels 6; Gaps 2;

QY 1 HDDEDDRGSHSIQGVQRCROERPRYSHARCVCQCRDQOQHGRHDEEGRGRMHG 60  
 Db 991 HEEQSHQSDRQTHVDEQNOQRQROTH---EENHD--HQGRHDEEHNHRQDHHQ 1044  
 QY 61 EGRE 65  
 Db 1045 QRRQ 1049

RESULT 10  
 ID PER\_DROME STANDARD; PRT; 1224 AA.  
 AC P07663; 017483; 024446; 024447; 024448; 024449;  
 DT 01-APR-1988 (Rel. 07, Created)  
 DT 15-DEC-1998 (Rel. 37, Last sequence update)  
 DT 15-DEC-1998 (Rel. 37, Last annotation update)  
 DE PERIOD CIRCADIAN PROTEIN (CLOCK-6 PROTEIN) (CLK-6).  
 GN PER.  
 OS Drosophila melanogaster (Fruit fly).  
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
 OC Ephydroidea; Drosophilidae; Drosophila.  
 RN [1]  
 RP SEQUENCE FROM N.A. (ISOFORMS PER-A/LONG; PER-B AND PER-C).  
 RC STRAIN-OREGON-R;  
 RL MEDLINE=87144607; PubMed=3102970;  
 RA Clitri Y., Colot H.V., Jacquier A.C., Yu Q., Hall J.C., Baltimore D.,  
 RA Rosbash M.;  
 RT "A family of unusually spliced biologically active transcripts encoded  
 RT by a Drosophila clock gene";  
 RL Nature 326:42-47(1987).  
 RN [2]  
 RP SEQUENCE FROM N.A. (ISOFORM PER-A/LONG).  
 RC STRAIN-CANTON-S;  
 RL Baylies M.K., Weiner L., Vossahl L.B., Saez L., Young M.W.;  
 RT "Genetic, molecular and cellular studies of the period locus and its  
 RT products in Drosophila melanogaster";  
 RL (in) Young M.W. (eds.);  
 RL Molecular genetics of biological rhythms, pp.123-153, Marcel Dekker,  
 RL New York (1993).  
 RN [3]  
 RP SEQUENCE FROM N.A. (ISOFORM PER-A/SHORT).  
 RC STRAIN-CANTON-S;  
 RL MEDLINE=86146900; PubMed=3081818;  
 RA Jackson F.R., Bargiello T.A., Yun S.-H., Young M.W.;  
 RT "Product of per locus of Drosophila shares homology with  
 RT proteoglycans";  
 RL Nature 320:185-188(1986).  
 RN [4]  
 RP SEQUENCE FROM N.A. (ISOFORMS PER-A/LONG; PER-A/SHORT; PER-D & PER-E).  
 RA Murphy L., Harris D., Barrell B.G.;  
 RL Submitted (NOV-1998) to the EMBL/GenBank/DBJ databases.  
 RN [5]  
 RP SEQUENCE OF 499-1075 FROM N.A. (ISOFORM PER-A).  
 RC STRAIN-OREGON-R;  
 RL MEDLINE=86245055; PubMed=3087625;  
 RA Reddy P., Jacquier A.C., Abovich N., Petersen G., Rosbash M.;  
 RT "The period clock locus of D. melanogaster codes for a proteoglycan";  
 RL Cell 46:53-61(1986).  
 RN [6]  
 RP SEQUENCE OF 63-573 FROM N.A. (ISOFORM PER-A).  
 RL MEDLINE=93170641; PubMed=8436278;  
 RA Kilman R.M., Hey J.;  
 RT "DNA sequence variation at the period locus within and among species  
 RT of the Drosophila melanogaster complex";  
 RL Genetics 133:375-387(1993).  
 RN [7]  
 RP PHOSPHORYLATION.  
 RX MEDLINE=94181572; PubMed=8134384;  
 RX Edery I., Zwiabel L.J., Dembinska M.E., Rosbash M.;  
 RA "Temporal phosphorylation of the Drosophila period protein.";

RL Proc. Natl. Acad. Sci. U.S.A. 91:2260-2264(1994).

CC -1- FUNCTION: ESSENTIAL FOR BIOLOGICAL CLOCK FUNCTIONS. DETERMINES THE PERIOD LENGTH OF CIRCADIAN AND ULTRADIAN RHYTHMS; AN INCREASE IN PER DOSAGE LEADS TO SHORTENED CIRCADIAN RHYTHMS AND A DECREASE LEADS TO LENGTHENED CIRCADIAN RHYTHMS. ESSENTIAL FOR THE CIRCADIAN RHYTHMICITY OF LOCOMOTOR ACTIVITY, ECLOSION BEHAVIOR, AND FOR THE RHYTHMIC COMPONENT OF THE MALE COURTSHIP SONG THAT ORIGINATES IN THE THORACIC NERVOUS SYSTEM. THE BIOLOGICAL CYCLE DEPENDS ON THE RHYTHMIC FORMATION AND NUCLEAR LOCALIZATION OF THE TIM-PER COMPLEX. LIGHT INDUCES THE DEGRADATION OF TIM, WHICH PROMOTES ELIMINATION OF PER. NUCLEAR ACTIVITY OF THE HETERODIMER COORDINATELY REGULATES PER AND TIM TRANSCRIPTION THROUGH A NEGATIVE FEEDBACK LOOP. BEHAVES AS A NEGATIVE ELEMENT IN CIRCADIAN TRANSCRIPTIONAL LOOP. DOES NOT APPEAR TO BIND DNA, SUGGESTING INDIRECT TRANSCRIPTIONAL INHIBITION.

CC -1- SUBUNIT: FORMS A HETERODIMER WITH TIMELESS (TIM); THE COMPLEX THEN TRANSLLOCATES INTO THE NUCLEUS.

CC -1- SUBCELLULAR LOCATION: NUCLEAR AT SPECIFIC PERIODS OF THE DAY. FIRST ACCUMULATES IN THE PERINUCLEAR REGION ABOUT ONE HOUR BEFORE TRANSLLOCATION INTO THE NUCLEUS. INTERACTION WITH TIM IS REQUIRED FOR NUCLEAR LOCALIZATION.

CC -1- ALTERNATIVE PRODUCTS: AT LEAST SIX ISOFORMS; PER-A/LONG (SHOWN HERE), PER-A/SHORT, PER-B, PER-C, PER-D AND PER-E; ARE PRODUCED BY ALTERNATIVE SPLICING.

CC -1- TISSUE SPECIFICITY: EXPRESSED IN NEURAL TISSUES AND IN SEVERAL NONNEURAL TISSUES OF THE ABDOMEN. MALPIGHIAN TUBULES CONTAIN A CIRCADIAN PACEMAKER THAT FUNCTIONS INDEPENDENTLY OF THE BRAIN. EXPRESSION OSCILLATES IN ALL TISSUES STUDIED EXCEPT FOR THE OVARY. PER-A ISOFORM IS MAINLY EXPRESSED IN ADULT'S HEAD.

CC -1- INDUCTION: EXPRESSION IS SENSITIVE TO TEMPERATURE BUT NOT TO LIGHT.

CC -1- DOMAIN: CONTAINS A REMARKABLE RUN OF ALTERNATING GLY-THR RESIDUES WHICH IS POLYMORPHIC IN LENGTH. AT LEAST THREE TYPES OF GLY-THR LENGTH EXIST IN THE NATURAL POPULATION, (GLY-THR)23 (SHOWN HERE), AND TWO MAJOR VARIANTS (GLY-THR)17 AND (GLY-THR)20. THIS GLY-THR STRETCH IS IMPLICATED IN THE MAINTENANCE OF CIRCADIAN PERIOD AT DIFFERENT TEMPERATURES. DELETION OF THE REPEAT LEADS TO A SHORTENING OF THE COURTSHIP SONG CYCLE PERIOD, AND THUS COULD BE IMPORTANT FOR DETERMINING FEATURES OF SPECIES-SPECIFIC MATING BEHAVIOR.

CC -1- DOMAIN: MUTATIONS IN THE PAS DOMAIN RESULT IN LONGER CIRCADIAN RHYTHMS AND COURTSHIP SONG (PERL MUTATION) OR MAKES THE FLIES ARRYTHMIC (PEROI MUTATION).

CC -1- PTM: PHOSPHORYLATED WITH A CIRCADIAN RHYTHMICITY, PROBABLY BY THE DOUBLE-TIME PROTEIN (DBT). PHOSPHORYLATION COULD BE IMPLICATED IN THE STABILITY OF PER MONOMER AND IN THE FORMATION OF HETERODIMER PER-TIM.

CC -1- SIMILARITY: TO OTHER G-T STRETCH CONTAINING PROTEINS.

CC -1- SIMILARITY: CONTAINS A PAS (PER-ARNT-SIM) DIMERIZATION DOMAIN.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC -----

DR EMBL: M30114: AAA28752.1: -

DR EMBL: M30114: AAA28753.1: -

DR EMBL: M30114: AAA28754.1: -

DR EMBL: AF033029: CAA27285.1: -

DR EMBL: AF033029: AAB87476.1: -

DR EMBL: M11969: AAA28751.1: -

DR EMBL: AL024485: CAA19677.1: -

DR EMBL: AL024485: CAA19678.1: -

DR EMBL: AL024485: CAA19679.1: -

DR EMBL: AL024485: CAA19680.1: -

DR EMBL: M13655: AAA28785.1: -

DR EMBL: D00009: BAA00007.1: -

DR EMBL: L07817: AAA28777.1: -

DR EMBL: L07818: AAA28776.1: -

DR EMBL: L07819: AAA28775.1: -

DR EMBL: L07821: AAA28773.1: -

DR EMBL: L07823: AAA28771.1: -

DR EMBL: L07825: AAA28769.1: -

DR PIR: A23932: UMF.

DR FLBASE: FB00003068: per.

DR INTERPRO: IPR000014: -

DR PFAM: PF00989: PAS; 2.

KW Biological rhythms; Repeat; Nuclear protein; Phosphorylation; NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).

CC POLYMORPHISM: Alternative splicing.

FT DOMAIN 66 79

FT REPEAT 227 295

FT REPEAT 378 445

FT DOMAIN 459 499

FT DOMAIN 694 743

FT DOMAIN 74 79

FT DOMAIN 143 146

FT DOMAIN 223 226

FT DOMAIN 872 879

FT DOMAIN 898 907

FT DOMAIN 906 914

FT DOMAIN 1006 1013

FT DOMAIN 1035 1041

FT DOMAIN 749 868

FT VARSPLIC 1 62

FT VARSPLIC 868 963

FT VARSPLIC 863 958

FT VARSPLIC 1076 1224

FT VARSPLIC 1155 1224

FT VARIANT 697 702

FT VARIANT 697 708

FT CONFLICT 211 211

FT CONFLICT 498 499

FT CONFLICT 637 637

FT CONFLICT 762 762

FT CONFLICT 762 762

FT CONFLICT 1029 1029

FT CONFLICT 1038 1038

FT CONFLICT 1075 1075

FT CONFLICT 1114 1114

FT CONFLICT 1215 1215

FT CONFLICT 1224 1224

FT SEQUENCE 1224 AA: 127852 MW: 71FA54ECF3E90FA4 CRC64;

Query Match 18.3%; Score 70; DB 1; Length 1224;

Best Local Similarity 36.2%; Pred. No. 5.3;

Matches 21; Conservative 9; Mismatches 14; Indels 14; Gaps 4;

QY 22 QEBPRYSHARCOVE--CRDDQ--OOHGRH-----BOEEQGRGWMHGEGEREE 66

DB 103 EEKPRRSIGCVGQICRELQDQHEDISEPAITIQLODEEEDDS--GSESDADIVE 159

RESULT 11

TRHY SHEEP

ID TRHY SHEEP STANDARD: PRT: 1549 AA.

AC P22793;

DT 01-AUG-1991 (rel. 19, Created)

DT 01-OCT-1994 (rel. 30, Last sequence update)

DT 01-OCT-2000 (rel. 40, Last annotation update)

DE TRICHOHYALIN.

GN THH.

OS Ovis aries (Sheep).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;



```

RT      "The gene for mouse epidermal filaggrin precursor. Its partial
RT      characterization, expression, and sequence of a repeating filaggrin
RT      unit.";
RN      J. Biol. Chem. 262:15643-15648(1987).
RN      (2)
RP      REVIEWS.
RA      Rochnagel J.A.;
RL      Submitted (SEP-1988) to the EMBL/GenBank/DBJ databases.
CC      -1- FUNCTION: AGGREGATES KERATIN INTERMEDIATE FILAMENTS AND PROMOTES
CC      DISULFIDE-BOND FORMATION AMONG THE INTERMEDIATE FILAMENTS DURING
CC      TERMINAL DIFFERENTIATION OF MAMMALIAN EPIDERMIS.
CC      -1- PPM: FILAGGRIN IS INITIALLY SYNTHESIZED AS A LARGE, INSOLUBLE,
CC      HIGHLY PHOSPHORYLATED PRECURSOR CONTAINING MANY TANDEM COPIES
CC      OF 248 AA, WHICH ARE NOT SEPARATED BY "LARGE LINERS". THE
CC      PRECURSOR IS DEPOSITED AS KERATOHYALIN GRANULES. DURING TERMINAL
CC      DIFFERENTIATION IT IS DEPHOSPHORYLATED & PROTEOLYTICALLY CLEAVED.
CC      -----
CC      THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on
CC      use by non-profit institutions as long as their content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (see http://www.isb-sib.ch/annouce/).
CC      or send an email to license@isb-sib.ch).
CC      -----
DR      EMBL; J03458; AAA75559.1; -.
DR      PIR; A28444; A28444.
DR      MGD; MG1:95553; FLG.
DR      Phosphorylation; Polypeptide; Developmental protein.
FW      NON_TER
FW      SEQUENCE 336 AA; 35678 MW; 259F124D3AC0DB2D CRC64;
SO
Query Match 18.2%; Score 69.5; DB 1: Length 336;
Best Local Similarity 27.7%; Pred. No. 1.7;
Matches 18; Conservative 12; Mismatches 26; Indels 9; Gaps 2;
OY 2 DDEDRRRGHSIQGCVQRQDERPRYSARQVQCRDDQDQGHNRHEEERG-GRG-RGMHG 60
DB 159 EESDSQHGHQHOHQBQGHQHOHQBQ-----EHEQPSGHRQDQSSGRGHQAHQ 210
OY 61 EGERE 65
DB 211 EGGRD 215
RESULT 13
PER_DROYA STANDARD; PRT; 1208 AA.
AC 024767; 026286;
DT 15-JUL-1999 (Rel. 38, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE PERIOD CIRCADIAN PROTEIN.
OS Drosophila yakuba (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachyceta; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
RN 11
RP SEQUENCE FROM N.A.
RA MEDLINE-91087242; Pubmed-1702156;
RA Thackeray J.R., Kyriacou C.P.;
RT "Molecular evolution in the Drosophila yakuba period locus.";
RL J. Mol. Evol. 31:389-401(1990).
RN 12
RP SEQUENCE OF 669-766 FROM N.A.
RA MEDLINE-93140158; Pubmed-1487825;
RA Peixoto A.A., Costa R., Wheeler D.A., Hall J.C., Kyriacou C.P.;
RT "Evolution of the threonine-glycine repeat region of the period gene
RT in the melanogaster species subgroup of Drosophila.";
RL J. Mol. Evol. 35:411-419(1992).
CC -1- FUNCTION: ESSENTIAL FOR BIOLOGICAL CLOCK FUNCTIONS. DETERMINES THE

```

CC PERIOD LENGTH OF CIRCADIAN AND ULTRADIAN RHYTHMS; AN INCREASE IN  
CC PER DOSAGE LEADS TO SHORTENED CIRCADIAN RHYTHMS AND A DECREASE IN  
CC LEADS TO LENGTHENED CIRCADIAN RHYTHMS. ESSENTIAL FOR THE CIRCADIAN  
CC RHYTHMICITY OF LOCOMOTOR ACTIVITY, ECLOSION BEHAVIOR, AND FOR THE  
CC RHYTHMIC COMPONENT OF THE MALE COURSHIP SONG THAT ORIGINATES IN  
CC THE THORACIC NERVOUS SYSTEM. THE BIOLOGICAL CYCLE DEPENDS ON THE  
CC RHYTHMIC FORMATION AND NUCLEAR LOCALIZATION OF THE TIM-PER  
CC COMPLEX. LIGHT INDUCES THE DEGRADATION OF TIM, WHICH PROMOTES  
CC ELIMINATION OF PER. NUCLEAR ACTIVITY OF THE HETERODIMER  
CC COORDINATIVELY REGULATES PER AND TIM TRANSCRIPTION THROUGH A  
CC NEGATIVE FEEDBACK LOOP. BEHAVES AS A NEGATIVE ELEMENT IN CIRCADIAN  
CC TRANSCRIPTIONAL LOOP. DOES NOT APPEAR TO BIND DNA, SUGGESTING  
CC INDIRECT TRANSCRIPTIONAL INHIBITION (BY SIMILARITY).  
CC  
CC -1- SUBUNIT: FORMS HETERODIMER WITH TIMELSS (TIM); THE COMPLEX THEN  
CC TRANSLOCATES INTO THE NUCLEUS (BY SIMILARITY).  
CC  
CC -1- SUBCELLULAR LOCATION: NUCLEAR AT SPECIFIC PERIODS OF THE DAY.  
CC FIRST ACCUMULATES IN THE PERINUCLEAR REGION ABOUT ONE HOUR BEFORE  
CC TRANSLOCATION INTO THE NUCLEUS. INTERACTION WITH TIM IS REQUIRED  
CC FOR NUCLEAR LOCALIZATION (BY SIMILARITY).  
CC  
CC -1- DOMAIN: THE RUN OF GLY-THR IS IMPLICATED IN THE MAINTENANCE OF  
CC CIRCADIAN PERIOD AT DIFFERENT TEMPERATURES. DELETION OF THE REPEAT  
CC LEADS TO A SHORTENING OF THE COURSHIP SONG CYCLE PERIOD, AND THUS  
CC COULD BE IMPORTANT FOR DETERMINING FEATURES OF SPECIES-SPECIFIC  
CC MATING BEHAVIOR (BY SIMILARITY).  
CC  
CC -1- PTM: PHOSPHORYLATED WITH A CIRCADIAN RHYTHMICITY, PROBABLY BY THE  
CC DOUBLE-TIME PROTEIN (DBT). PHOSPHORYLATION COULD BE IMPLICATED IN  
CC THE STABILITY OF PER MONOMER AND IN THE FORMATION OF HETERODIMER  
CC PER-TIM (BY SIMILARITY).  
CC  
CC -1- SIMILARITY: TO OTHER G-T STRETCH CONTAINING PROTEINS.  
CC  
CC -1- SIMILARITY: CONTAINS A PAS (PER-ARMY-SIM) DIMERIZATION DOMAIN.  
CC  
CC  
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC DR EMBL; X61127; CAA43439.1; -.  
CC DR EMBL; S53298; AAB25029.1; -.  
CC DR FLYBASE; FBgn0013215; Dya4\per.  
CC DR INTERPRO; IPR000014; -.  
CC DR INTERPRO; IPR001014; -.  
CC DR PFAM; PF00989; PAS; 2.  
CC DR PRINTS; PRO0308; ANTIREEZEI.  
CC KW Biological rhythms; Repeat; Nuclear protein; Phosphorylation.  
CC FT DOMAIN 66 79  
CC FT REPEAT 231 299  
CC FT REPEAT 382 449  
CC FT DOMAIN 463 503  
CC FT DOMAIN 698 741  
CC FT DOMAIN 74 79  
CC FT DOMAIN 857 864  
CC FT DOMAIN 889 893  
CC FT DOMAIN 987 997  
CC FT DOMAIN 1019 1025  
CC FT DOMAIN 1019 1025  
CC FT CONFLICT 742 742  
CC FT CONFLICT 751 751  
CC FT CONFLICT 762 762  
CC FT CONFLICT 762 762  
CC SEQUENCE 1208 AA; 126820 MW; E534A35E473B8483 CRC64;

```

Query Match      18.2% Score 69.5; DB 1; Length 1208;
Best Local Similarity 34.5%; Pred. No. 5.6;
Matches 20; Conservative 6; Mismatches 19; Indels 13; Gaps 3;

QY      22 QERYRYSHARC---VOECRD--DOOGHGRHEOE-----EEOGRGRRGWHGEGERE 66
          :|::||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::
Db       104 EERRPSSGGGVGEQSCRELLDDPOQHGDHSSEKATIEQLQEGEDSGSSEAEARVE 161

RESULT 14
```

```

ID PER_DROSE STANDARD: PRT: 661 AA.
AC 000354; Q26285;
DT 01-OCT-1993 (Rel. 27, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE PERIOD CIRCADIAN PROTEIN (FRAGMENTS).
OS
GN Drosophila sechellia (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydriidae; Drosophilidae; Drosophila.
RN [1]
RP SEQUENCE OF 1-558 FROM N.A.
RX MEDLINE=93170641; Pubmed=8436278;
RA Kilman R.M., Hey J.;
RT "DNA sequence variation at the period locus within and among species
RT of the Drosophila melanogaster complex.";
RL Genetics 133:375-387(1993).
RN [2]
RP SEQUENCE OF 559-661 FROM N.A.
RX MEDLINE=93140158; Pubmed=1487825;
RA Pelxoto A.A., Costa R., Wheeler D.A., Hall J.C., Kyriacou C.P.;
RT "Evolution of the threonine-glycine repeat region of the period gene
RT in the melanogaster species subgroup of Drosophila.";
J. Mol. Evol. 35:411-419(1992).
CC -1- FUNCTION: ESSENTIAL FOR BIOLOGICAL CLOCK FUNCTIONS. DETERMINES THE
CC PERIOD LENGTH OF CIRCADIAN AND ULTRADIAN RHYTHMS; AN INCREASE IN
CC PER DOSAGE LEADS TO SHORTENED CIRCADIAN RHYTHMS AND A DECREASE
CC LEADS TO LENGTHENED CIRCADIAN RHYTHMS. ESSENTIAL FOR THE CIRCADIAN
CC RHYTHMICITY OF LOCOMOTOR ACTIVITY, ECLOSION BEHAVIOR, AND FOR THE
CC RHYTHMIC COMPONENT OF THE MALE COURTSHIP SONG THAT ORIGINATES IN
CC THE THORACIC NEUVOUS SYSTEM. THE BIOLOGICAL CYCLE DEPENDS ON THE
CC RHYTHMIC FORMATION AND NUCLEAR LOCALIZATION OF THE TIM-PER
CC COMPLEX. LIGHT INDUCES THE DEGRADATION OF TIM, WHICH PROMOTES
CC ELIMINATION OF PER. NUCLEAR ACTIVITY OF THE HETERODIMER
CC COORDINATIVELY REGULATES PER AND TIM TRANSCRIPTION THROUGH A
CC NEGATIVE FEEDBACK LOOP. BEHAVES AS A NEGATIVE ELEMENT IN CIRCADIAN
CC TRANSCRIPTIONAL LOOP. DOES NOT APPEAR TO BIND DNA, SUGGESTING
CC INDIRECT TRANSCRIPTIONAL INHIBITION (BY SIMILARITY).
CC -1- SUBUNIT: FORMS HETERODIMER WITH TIMELESS (TIM); THE COMPLEX THEN
CC TRANSLOCATES INTO THE NUCLEUS (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: NUCLEAR AT SPECIFIC PERIODS OF THE DAY.
CC FIRST ACCUMULATES IN THE PERINUCLEAR REGION ABOUT ONE HOUR BEFORE
CC TRANSLOCATION INTO THE NUCLEUS. INTERACTION WITH TIM IS REQUIRED
CC FOR NUCLEAR LOCALIZATION (BY SIMILARITY).
CC -1- DOMAIN: THE RUN OF GLY-THR IS IMPLICATED IN THE MAINTENANCE OF
CC CIRCADIAN PERIOD AT DIFFERENT TEMPERATURES. DELETION OF THE REPEAT
CC LEADS TO A SHORTENING OF THE COURTSHIP SONG CYCLE PERIOD, AND THUS
CC COULD BE IMPORTANT FOR DETERMINING FEATURES OF SPECIES-SPECIFIC
CC MATING BEHAVIOR (BY SIMILARITY).
CC -1- PTM: PHOSPHORYLATED WITH A CIRCADIAN RHYTHMICITY, PROBABLY BY THE
CC DOUBLE-TIME PROTEIN (DBT). PHOSPHORYLATION COULD BE IMPLICATED IN
CC THE STABILITY OF PER MONOMER AND IN THE FORMATION OF HETERODIMER
CC PER-TIM (BY SIMILARITY).
CC -1- SIMILARITY: CONTAINS A PAS (PER-ARNT-SIM) DIMERIZATION DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation.
CC The European Bioinformatics Institute. There are no restrictions on
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announcement/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; L07820; AAA28774.1; -
DR EMBL; L07822; AAA28772.1; -
DR EMBL; L07824; AAA28770.1; -
DR EMBL; L07827; AAA28767.1; -
DR EMBL; S53297; AAB25028.1; -
DR FLYBASE: FBgn0012799; Dsec\per.
DR PFM: PF00989; PAS; 2.
TW Biological rhythms; Repeat; Nuclear protein; Phosphorylation.

```

```

FT      NON_TER          1           66         NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).
FT      DOMAIN            53           66
FT      DOMAIN            61           66
FT      DOMAIN            208          211       POLY-LYS.
FT      REPEAT            212          280       POLY-ALA.
FT      REPEAT            363          430       PAS-1.
FT      DOMAIN            444          484       PAS-2.
FT      VARIANT            85           85       PAC MOTIF.
FT      NON_CONS          558          559       L-> M (IN CLONE SE-P4) .
FT      DOMAIN            591          631
FT      NON_TER            661
SQ      SEQUENCE          661 AA;   69677 MW;   C0680E5198B6A209 CRC64;

Query Match               18.1%, Score 69; DB 1; Length 661;
Best Local Similarity     36.2%; Pred. No. matches 7;
Matches 21; Conservative 8; Misorderings 13; Indels 16; Gaps 4;

Oy    22 QERPVSIMARCGQE--CRDDQ-QQHGRH-----EQEEGRCRGTHGEGRRE 66
Db    90 EKKPPSGTCVEQQICRHLDPDGHGDHSFPOAAELQLDEEDDS--GSSEADRV 144
        ::::| |::: ||::| ||||| :|||:| | | | |

RESULT 15
CIRK_MOUSE
ID CIR4_MOUSE STANDARD: PRT; 654 AA.
AC Q61423;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE VOLTAGE-GATED POTASSIUM CHANNEL PROTEIN KVL1.4.
GN KCNA4.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
CC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=AKR;
RX MEDLINE=94292198; PubMed=8020965;
RA Wymore R.S., Korenberg J.R., Kinoshita K.D., Ayar J., Coyne C.,
RA Chen X.N., Husted C.M., Copeland N.G., Gutman G.A., Jenkins N.A.,
RA Chandy K.G.;
RT "Genomic organization, nucleotide sequence, biophysical properties,
RT and localization of the voltage-gated K+ channel gene KCNA4/Kv1.4 to
RT mouse chromosome 2/human 11p14 and mapping of KCNC1/KV3.1 to mouse
RT 7/human 11p14.3-p15.2 and KCNA1/Kv1.1 to human 12p13." ;
RL Genomics 20:191-202(1994).
CC -!- FUNCTION: THIS PROTEIN MEDIATES THE VOLTAGE-DEPENDENT POTASSIUM
CC ION PERMEABILITY OF EXCITABLE MEMBRANES. ASSUMING OPENED OR CLOSED
CC CONFORMATIONS IN RESPONSE TO THE VOLTAGE DIFFERENCE ACROSS THE
CC MEMBRANE, THE PROTEIN FORMS A POTASSIUM-SELECTIVE CHANNEL THROUGH
CC WHICH K+ IONS MAY PASS IN ACCORDANCE WITH THEIR ELECTROCHEMICAL
CC GRADIENT.
CC -!- SUBUNIT: HETEROTETRAMER OF POTASSIUM CHANNEL PROTEINS (PROBABLE).
CC -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
CC -!- DOMAIN: THE AMINO TERMINUS MAY BE IMPORTANT IN DETERMINING THE
CC RATE OF INACTIVATION OF THE CHANNEL WHILE THE TAIL MAY PLAY A ROLE
CC IN MODULATION OF CHANNEL ACTIVITY AND/OR TARGETING OF THE CHANNEL
CC TO SPECIFIC SUBCELLULAR COMPARTMENTS.
CC -!- DOMAIN: THE SEGMENT S4 IS PROBABLY THE VOLTAGE-SENSOR AND IS
CC CHARACTERIZED BY A SERIES OF POSITIVELY CHARGED AMINO ACIDS AT
CC EVERY THIRD POSITION.
CC -! SIMILARITY: THIS CHANNEL PROTEIN BELONGS TO THE A-TYPE POTASSIUM
CC CURRENT CLASS. BELONGS TO SHAKER POTASSIUM CHANNEL SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation at
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to licensese@isb-sib.ch).
```

DR EMBL: U03723: AAB60668.1; -.  
DR HSSP: P15385; 1ZTO.  
DR MGD: MGI:96661; KCNA4.  
DR INTERPRO: IPR000636; -.  
DR INTERPRO: IPR003091; -.  
DR PFAM: PF00520: ion.trans; 1.  
DR PRINTS: PR00169; KCHANNEL.  
KW Ionic channel; Transmembrane; Ion transport; Voltage-gated channel;  
KW Glycoprotein; Multigene family; Phosphorylation.  
FT TRANSMEM 309 327 SEGMENT S1 (POTENTIAL).  
FT TRANSMEM 372 393 SEGMENT S2 (POTENTIAL).  
FT TRANSMEM 405 425 SEGMENT S3 (POTENTIAL).  
FT TRANSMEM 444 462 SEGMENT S4 (POTENTIAL).  
FT TRANSMEM 479 498 SEGMENT S5 (POTENTIAL).  
FT TRANSMEM 540 562 SEGMENT S6 (POTENTIAL).  
FT DOMAIN 38 50 POLY-ALA.  
FT DOMAIN 62 65 POLY-HIS.  
FT DOMAIN 83 89 POLY-ARG.  
FT DOMAIN 123 137 POLY-GLU.  
FT DOMAIN 162 166 POLY-GLY.  
FT DOMAIN 434 437 POLY-GLN.  
FT CARBOHYD 182 182 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 353 353 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 643 643 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT MOD\_RES 600 600 PHOSPHORYLATION (BY CAPK) (POTENTIAL).  
SQ SEQUENCE 654 AA; 73473 MW; 8693651A30BD36D4 CRC64;

Query Match 17.9%; Score 68.5; DB 1; Length 654;  
Best Local Similarity 27.3%; Pred. No. 4.1;  
Matches 15; Conservative 10; Mismatches 27; Indels 3; Gaps 1;  
OY 11 HSLQOCVORCRQERPRYSARCVQECRDDQOHGRHEQEEQGR--GRGMHGEQ 62  
Db 96 HHROSFPHCSDLMPGSGSEKILRELSEEEDEDEEEDEEEEGRFFYYSEEDHDG 150

Search completed: March 1, 2001, 16:20:34  
Job time: 209 sec